DOCUMENT RESUME

ED 128 370 TM 005 492

AUTHOR McDonald, Frederick J.
TITLE Designing Research for Policy Making.
PUB DATE [Apr 76]
NOTE 36p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Academic Achievement; *Decision Making; *Educational Policy; *Educational Research; Effective Teaching; Elementary Education; Mathematics; *Policy Formation; Reading; *Research Design; Teacher Behavior

ABSTRACT

Phase II of the Beginning Teacher Evaluation Study was designed to provide data relevant to policies that might be developed for teacher preparation and licensing. A portion of the data from Phase II is used to illustrate how the research was designed to answer these policy questions and how the data implies provisional answers to them. To answer the question of what is awry between the world of policy and the world of research, two answers are proposed. The first is that the failure to understand the political nature of policy making leads to irrelevant research which offers policy makers solutions they cannot use. The policy alternatives available to the policy makers should be carefully laid out as a first step in planning research. The second answer is that since policy making is decision making, the decision making process also has to be laid out to see what kind of decisions need to be made to develop and implement a policy. An analysis of these decisions should indicate the kinds of information that policy makers will need. These needs should determine which research needs to be done. (RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED128370

DESIGNING RESEARCH FOR POLICY MAKING[1]

Frederick J. McDonald
Educational Testing Service
Princeton, New Jersey

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

Frederick J. McDonald

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Paper presented at the Annual Meeting of the
American Educational Research Association
San Francisco, California
April, 1976

[1]Data cited in this paper were selected from the data gathered in Phase II of the Beginning Teacher Evaluation Study. This phase of the research was done by Educational Testing Service for the California Commission for Teacher Preparation and Licensing and funded by the National Institute of Education.

Copyright © 1976 by Educational Testing Service. All rights reserved.

TM005 492

Preface

Phase II of the Beginning Teacher Evaluation Study was designed to provide data relevant to policies that might be developed for teacher preparation and licensing. I have used a portion of the data from Phase II to illustrate how the research was designed to answer these policy questions and how the data implies provisional answers to them.

To illustrate how the research helps to define or to sharpen policy issues, I must speculate about the implications of the data for policy. These are my own speculations. They have not been made as recommendations to the Commission, nor are they all the implications for policy that may be drawn from the Phase II data. I attempt only to illustrate how such data may be used to begin to shape policies. I do so because I think that it is important that such speculations be made; that as research develops, its probable implications for policy be anticipated; and that at each stage of a long-term research project, cognizance be taken of its implications for policy development.

The usual practice of researchers is to avoid drawing implications for policy from research. I think that this practice is short-sighted and self-defeating because it leads policy-makers to believe that research has little to say of relevance to policy. It seems to me that a reasonable position is to suggest which policies are consistent with the data and which are not, what might be done and what ought not to be done.

Since Phase II was designed to illuminate certain policy questions, it is appropriate to state whether illumination has resulted. That is what I propose to do in what follows.

INTRODUCTION

The belief that educational research should inform educational policy-making seems to be widely held among educators and laymen alike, and of course, by educational researchers. It is a benign belief since it is rarely acted upon. It is also a fortuitous one because it stimulates periodic calls for more research, better research, different research, and has in the recent past produced generous funding for research.

But this faith when transformed into action has frequently bred disillusionment. Research programs produce meager or confusing or even contradictory results. Researchers themselves frequently point out how little they know and warn policy-makers to use even that little with caution or not at all. Policy-makers find no easy or quick solutions in research; and many have come to believe that research has little practical import.

What is awry between the world of policy and the world of research? The customary answers to this question are trite. I propose here two other answers. The first is that the failure to understand the political nature of policy-making leads to irrelevant research which offers policy-makers solutions they cannot use.

The obvious fact about policy-makers is that they are not free to make whatever policy seems good to them. They are elected or appointed, and are accountable to their constituencies or appointer. They are part of an institution or agency with prescribed and limited powers. There are things they can do and cannot do.

My view is that the policy alternatives available to the policy-makers should be carefully laid out as a first step in planning research. It should be obvious that one cannot know which previous research may be relevant or which research design will be appropriate if one does not know what actions a policy-making group may consider.

The second answer that I propose is that since policy-making is decision-making, the decision-making process also has to be laid out to see what kinds of decisions need to be made to develop and implement a policy. An analysis of these decisions should indicate the kinds of information the policy-makers will need. These needs should determine which research needs to be done.

By way of contrast, consider the usual approach when policy-makers and researchers come together. The researcher is asked, for example, what do you know about teacher effectiveness? Since the researcher's caution usually exceeds his or her knowledge, the answer is likely to be tentative at best. The policy-makers conclude that research has little to offer. Little to offer for what? What do the policy-makers want to do? What can they do? What costs can they tolerate to achieve their goals? What decisions do they have to make to achieve their goals? Do they want to find out whether one alternative is better than another? Do they want other alternatives to consider? Do they want to find out the consequences of implementing an alternative they have in mind? Do they want to compare the costs of several alternatives? These questions frame the research design likely to produce results useful to the policy-makers.

The policy-makers should first describe the goals to be achieved by the policy, the alternatives they can and will consider, and the resources they can use to implement the policy. Research useful to policy-making may then be designed to answer questions such as:

1. What are the effects of a policy alternative likely to be?
2. Which alternatives are more likely to achieve the intended goals of a policy?
3. Are these alternatives likely to have undesirable effects?
4. What are the benefits and costs of each alternative?

The customary practice of implementing a policy and then asking the researchers to evaluate its effects is disastrous as much of recent evaluation research illustrates. This practice rushes a seemingly good idea into a program without considering if there are equally good alternatives or without evaluating the costs of the alternatives.

The Beginning Teacher Evaluation Study has departed from this way of making and evaluating policy. The California Commission on Teacher Preparation and Licensing chose wisely to postpone policy-making until it had sufficient data on which to base policy. But what kind of research should be done even in this ideal situation? Before answering that question, a more detailed analysis will be useful of how policy may be set and how research may be relevant to the formation of policy.

WHAT IS POLICY-MAKING?

A policy is always aimed at taking action. A policy may take one or all of several forms. It may be a statement of a principle: e.g., "teacher education programs will be a mixture of theoretical and practical courses." It may be a prescription of a definite course of action which is to guide present and future decisions: "admission to a teacher education program will be based on demonstrated intellectual competence, appropriate academic preparation, and demonstrated ability to work effectively with children." The first policy leaves decisions about the appropriate mixture of theoretical and practical courses to the persons who will carry out the policy. It excludes programs solely theoretical or practical. The second policy stipulates that there will be an admissions procedure which will utilize certain kinds of information about applicants but leaves to the admissions committee the methods by which the information will be gathered and the criteria for selection.

A third kind of policy specifies both what is to be done and how to do it: "admission to a teacher education program will be based on evidence of demonstrated capacity to work with children; such evidence shall be extended work experience in such facilities as day-care centers, nursery schools, recreational programs, and tutoring programs; evidence of such work will be accompanied by evaluations by the applicant's supervisor or employer." This policy states the requirement specifically both as to its substance and evidence of its fulfillment. It leaves only the meaning of the term, "extended," to the discretion of the admissions committee.

A fourth kind of policy states an objective and a program of action to achieve that objective: "a major goal of practical training for teaching is to learn the skills of managing groups of children; to achieve this goal each teacher education program in the state will provide specific training programs in classroom management, supervised practice in conducting instruction, and will evaluate each trainee's skill in managing instruction." This policy has stipulated an objective and the kind of program by which it is to be achieved. It leaves the details of the program to the discretion of the teacher educators.

Only one of these policies states an objective explicitly but the unstated objectives are easily inferred. In one case (the first policy) the actions to be taken are left to the discretion of the teacher educators but the principle of action is indicated. In another case (the fourth policy) specific actions to be taken are prescribed. All of these policies have in common the assumption that their implied or prescribed actions will lead to their objectives.

The art of policy-making lies in selecting objectives which are likely to be achieved if the prescribed actions are taken. If these actions do not achieve their intended goal, the policy is regarded as ineffectual; if they do, it is a success. The fallibility of policy-makers lies in their assumptions about cause and effect.

We expect policy-makers to be clear about the goals to be achieved. We expect them to make wise choices among alternative ways of achieving these goals. But neither we nor they may have sufficient knowledge to be reasonably certain that a choice is wise.

Consider the case of the California Commission for Teacher Preparation and Licensing. Their goal is to insure a supply of competent teachers. They can set standards for selecting, training and evaluating prospective teachers. They can specify the content of programs of training. What alternatives are available to them? Should they set standards for what teachers should know or be able to do or both? How much does a teacher's effectiveness depend on his or her attitudes towards teaching and children? If it does, should attitudes be considered in selecting prospective teachers; should their attitudes be evaluated before licensing them? How much is this effectiveness related to the kinds of children taught? If it is, should teachers be licensed to teach the kinds of children with whom they have been effective or should minimal common standards be set for all teachers? Knowledge about the relation to learning of each of these alternatives or the variables underlying them and about their relation to each other is needed to make wise policy choices.

One contribution of research to policy-making is facts about the effectiveness of policy alternatives. But should the research be designed to compare the effectiveness of the alternatives which the policy-makers propose or should it also be used to generate alternatives which the policy-makers may not have considered? The latter course seems more desirable.

The appearance of rationality in this conclusion is an illusion created by the deceptive simplicity of such phrases as, "alternatives that the policy-makers will consider; or, "alternatives that the research may generate." This

language implies that policy-makers lay out their alternatives like proper thinking machines for analysis by research methods. Or, it implies that policy-makers will seriously consider alternatives that research may develop.

A common belief is that once research has done its work, policies will be obvious. If, for example, research shows a high positive correlation between a teaching performance and learning, then it should follow that teachers should be trained on the performance. Such "applications" of research rarely occur because policy-makers must consider many factors in addition to the research evidence. The feasibility of the policy is a prime consideration. What can be done, what will it cost to do it, and how great will the benefits be? These are important questions whose answers tell the policy-makers if they have a workable policy.

To be feasible a policy must meet three criteria. The policy-makers must have the power to enact the policy. The policy must be accepted by those who will enact it. Its costs must be reasonable.

A group of policy-makers must have the specific power to promulgate a policy. The California Commission for Teacher Preparation and Licensing has the power to set standards for preparation and certification. It does not have the power to reorganize the functions and role of the teacher or to prescribe the content of the curriculum of the schools. If curricula, or school organization, or the social background of the pupils are more significant influences on what and how much children learn than the skill of teachers, the Commission through its policies cannot directly affect these aspects of schooling.

The goal of the research on teaching effectiveness should be to determine the contribution of teaching skill to learning and also to determine how teaching skill combines with specific programs, or is affected by organization of the teaching staff, or by the administrative style of the principal, or by the characteristics of the pupils being taught. The policy-makers should be able to conclude from this research whether the alternative they may enact will have a significant effect on learning.

A second determinant of a policy's feasibility is its cost. A policy will be difficult to enact if its costs are not reimbursed when it is enacted. Two other kinds of costs should be considered but in educational research rarely are. These are marginal costs and opportunity costs.

The marginal cost is the increase in the total cost needed to achieve an increment in a desired outcome. The desired outcome in this case is the increase in pupil learning which results from improved teacher performance or knowledge or attitude. The marginal cost is the increase in the total cost of selecting, training, and evaluating a teacher to increase their effectiveness. Such costs can be estimated if we know the regression equation which relates the teaching variables to learning, and if we know the costs of selecting, or training, or evaluating teachers. Research can provide this information.

Opportunity cost is the cost of alternatives foregone. What could have been done with the money spent on the additional training? The alternative uses of the money are assumed to be beneficial. Obviously the benefits to be achieved by the additional training must outweigh the opportunity costs.

This description of the potential costs of a policy-innovation invites the conclusion that a cost-benefit analysis is an integral part of policy-making. Research methods are available for making this type of an analysis.

A third determinant of feasibility and one difficult to estimate is the acceptability of a proposed policy. Some policies will be undermined or fulfilled only in appearance if they are not accepted by those who must enact them. Some policies will be openly opposed.

A common delusion is that research data supporting a policy will convince people of the policy's wisdom. Obviously the reasons for opposing a policy are many, and some so personally important that empirical evidence will be ignored, or, as is more likely, attacked as inadequate.

The policy-makers need data on who will oppose or undermine a policy, why they will oppose it, and what they will find acceptable. Research methods are available to provide this information to policy-makers who can use it to shape a policy or to stage its implementation.

A feasible policy can be developed if these three aspects of policy-making are carefully studied in each instance that a policy is being shaped. A research program to assist in this decision-making has three phases each directed to a different goal. The first phase is directed to finding out if teaching performance or knowledge or some other aspect of teaching makes a difference in pupil learning and how much of a difference they make. It also determines if there are interactions among the teaching variables and other aspects of schooling. This research is the groundwork for identifying policy-alternatives and their probable effectiveness. The

second phase studies how by selecting or training to improve those characteristics of teachers found to be related significantly to learning. This research should also be used to estimate the costs of selecting or training. The third phase is the cost-benefits analysis of various alternatives. The fourth phase studies the conditions under which a new policy is likely to be accepted and enacted.

A research program of this kind provides data for the decisions which the policy-makers have to make to develop a useful policy. This plan for research proposes that different research methods be used to answer the questions which are fundamental to making these decisions. Research is used as a way of solving a problem. The problem is to develop policies which will guarantee a supply of competent teachers whose benefits outweigh its costs; which will be accepted by teacher educators and teachers' organizations; and which fall within the purview of the Commission's delegated powers.

PHASE II AND POLICY ISSUES

In the earliest days of Phase II the Commission indicated that it wanted information that would help it shape policies for training and licensing. They indicated the broad categories of alternatives they could consider: selection procedures, course requirements, evaluation procedures, specific training requirements. The study was designed accordingly to provide data relevant to a variety of questions arising from these possibilities for making policy.

The focus of the study was the relation of teaching performances to pupil achievement in reading and mathematics. The areas of achievement in these subjects were to be in reading: comprehension skill, decoding skills, application skills and attitudes towards reading; and in mathematics: concept comprehension, computational skill, applications skills, and attitudes towards mathematics. A very large number of measures were then used to assess teachers' aptitudes, knowledge, attitudes and background; their pupils' aptitudes and background; and the characteristics of the teaching environment.

The analysis of the data had two major components. Each of these components will be the basis for a discussion of how the results of Phase II may be used to clarify, sharpen and define policy.

In this discussion I will confine the report to the analyses of the teacher variables. I will make no attempt to summarize all the results of the study but will select some of them to illustrate how this phase of the Beginning Teacher Evaluation Study contributes to the development of policy.

Two analyses are required to evaluate the comparative effects of the different teacher variables on learning. The first step is to analyze the effects of teaching performances; the second step is to see how the other variables are related to teaching performances.

To assess the effects of teaching performances, assessments of pupils' skills in reading and mathematics were made in the fall and again in the spring. Teachers' classroom performances were observed in the intervening period and

the teachers reported twice on their teaching goals, content and methods. These data were used in multiple and stepwise regressions to estimate the relation of teaching performance to the different measures of pupils' achievement. Both residual and mean change scores were used in these analyses. One set of results from the analysis of second-grade reading is used to illustrate what was found that has implications for policy-making.

Table 1 summarizes the results of regressing mean-difference scores in second-grade reading on the teaching performance variables. Teaching performances were significantly related to decoding scores at the .05 significance level and to one of the measures of reading comprehension at the .09 significance level.

Table 2 presents the results of regressing mean-difference scores in decoding on the performance variables. $R = .8704$ and $R^2 = .7575$. The teaching performance variables account for almost three-fourths of the variance in the mean differences in decoding scores. Seven of the variables used in this analysis have significant regression weights; four are positive predictors and four are negative:

+	-
WD-1: Teacher Instructional Time	WD-5: Variety of Instructional Materials
AP-7: Teaching the Class-as-a-whole	R-2 : Amount of New Content Introduced Each Lesson
AP-9: Questions Asked and Answered	AP-11: Teacher Location
AP-12: Pattern of Corrective Feedback	AP-13: Controlling Attention

Table 1

R's, R^2, F-Values, and Significance Levels Obtained in Five Multiple-Regression Analyses of Mean-Difference Scores in Reading on 22 Teaching Performance Variables; Grade 2 Reading (N = 39)

	R	R^2	F	p
CAT - COMPREHENSION	.8435	.7115	1.9054	.0894
READING APPLICATION	.6571	.6571	.5871	.8805
DECODING	.8704	.7574	2.4140	.0341
READING ACHIEVEMENT	.8127	.6604	1.5029	.1975
TOTAL READING	.6582	.4332	.5905	.8779

THE DEPENDENT VARIABLE IS DECODING. THE MULTIPLE CORRELATION IS 0.670+. THE STANDARD ERROR OF ESTIMATE= 0.114

	SUM OF SQUARES	PROPORTION OF SQUARES	D.F.	MEAN SQUARE	F RATIO	PROBABILITY OF LARGER F
TOTAL ABOUT ORIGIN	1.0000		39.			
UNDER NULL HYPOTHESIS	1.0000	1.0000	39.			
DUE TO HYPOTHESIS	0.7575	0.7575	22.	0.0344	2.4140	0.0041
ERROR	0.2425	0.2425	17.	0.0143		

	STANDARD REG. WEIGHT	REGRESSION WEIGHTS	STANDARD ERROR OF WT.	T STATISTICS WITH 17 D.F.	CONTRIBUTION TO R-SQ.	MEASURE OF COLLINEARITY
INDEPENDENT VARIABLES						
AB-----1	0.6838	0.6838	0.2345	2.9156	-0.1213	0.7407
AB-----2	0.3247	0.3247	0.2021	1.6021	-0.0360	0.6526
AB-----3	-0.2490	-0.2490	0.1903	-1.3087	-0.0244	0.6060
AB-----4	0.2594	0.2594	0.1912	1.3568	-0.0263	0.6096
AB-----5	-0.6839	-0.6839	0.2471	-2.8844	-0.1187	0.7405
RAMOS--1	0.3284	0.3284	0.2423	1.3557	-0.0258	0.7575
RAMOS--2	-0.6608	-0.6608	0.2417	-2.5478	-0.0926	0.7416
RAMOS--3	0.4391	0.4391	0.2452	1.7905	-0.0457	0.7628
RAMOS--4	0.0928	0.0928	0.2131	0.4355	-0.0027	0.6858
APPLE--1	0.1095	0.1095	0.1858	0.5892	-0.0050	0.5869
APPLE--2	0.0497	0.0497	0.2418	0.2055	-0.0006	0.7560
APPLE--3	-0.6820	-0.6820	0.3058	-2.2303	-0.0709	0.8474
APPLE--5	0.2547	0.2547	0.1697	1.5015	-0.0328	0.5045
APPLE--6	0.0493	0.0493	0.2193	0.2291	-0.0007	0.6924
APPLE--7	0.8370	0.8370	0.2764	3.0287	-0.1333	0.8096
APPLE--8	0.2528	0.2528	0.1619	1.5610	-0.0340	0.4560
APPLE--9	0.4393	0.4393	0.2190	2.0065	-0.0574	0.7025
APPLE-10	0.1616	0.1616	0.1908	0.8469	-0.0102	0.6064
APPLE-11	-0.3948	-0.3948	0.1971	-1.9521	-0.0544	0.6330
APPLE-12	0.6770	0.6770	0.2217	3.0533	-0.1330	0.7099
APPLE-13	-0.5832	-0.5832	0.2626	-2.1447	-0.0654	0.7532
APPLE-14	-0.1505	-0.1505	0.1824	-0.8275	-0.0058	0.5741

Table 2

The Results of the Multiple Regression of Mean-Change Decoding Scores on Teacher Performance Scores - Grade 2 (N_T = 39)

The positive predictors suggest a style in which the teacher instructs the class-as-a-whole, interacts frequently by questioning and giving feedback; and circulates about the room. Visualize the following scene: the class has reading materials available (on the average children spend 50 percent of the instructional time in seatwork); the teacher gives an explanation of the reading task, circulates among the pupils, and asks and answers questions, giving corrective feedback. The results suggest that a teacher who uses this strategy consistently produces on-the-average greater gains in decoding skills.

The negative predictors suggest a style of teaching performance which is dysfunctional: using a variety of reading materials, spending more time keeping pupils on-task, introducing more new content per lesson, and not circulating among the pupils. Again visualize the picture of pupils working at their desks, but now the teacher is not circulating among them, not asking and answering questions, giving little feedback, and spending more time keeping pupils on-task than instructing.

Now consider the regression analysis in which the significance level was .09. In this analysis the outcome measure was reading comprehension as measured by the California Reading Achievement Test. The results of this analysis are in Table 3. In this analysis, there were two significant positive predictors (both of these are different measures of the same variable).

+	-
WD-5: Variety of Instructional Materials	WD-1: Teacher Instructional Time
R -3: Variety of Instructional Materials	

THE DEPENDENT VARIABLE IS CAT RCMP. THE MULTIPLE CORRELATION IS 0.8[illegible] THE STANDARD ERROR OF ESTIMATE 0.1303

	SUM OF SQUARES	PROPORTION OF SQUARES	D.F.	MEAN SQUARE	F RATIO	PROBABILITY OF LARGER F
TOTAL ABOUT ORIGIN UNDER NULL HYPOTHESIS	1.0000		39.			
DUE TO HYPOTHESIS	1.0000	1.0000	39.			
ERROR	0.7115	0.7115	22.	0.0323	1.9054	0.0694
	0.2885	0.2885	17.	0.0170		

INDEPENDENT VARIABLES	STANDARD REG. WEIGHT	REGRESSION WEIGHTS	STANDARD ERROR OF WT.	T STATISTICS WITH 17 D.F.	CONTRIBUTION TO R-SQ.	MEASURE OF COLLINEARITY
MC-----1	-0.5352	-0.5352	0.2558	-2.0920	-0.0743	0.7407
MC-----2	-0.1331	-0.1331	0.2211	-0.6[illegible]	-0.0061	0.6528
MC-----3	-0.0169	-0.0169	0.2078	-0.0814	-0.0001	0.6030
MC-----4	0.1023	0.1023	0.2080	0.4907	-0.0041	0.6095
MC-----5	0.6218	0.6218	0.2586	2.4046	-0.0581	0.7463
RAMOS--1	-0.4004	-0.4004	0.2646	-1.5134	-0.0389	0.7575
RAMOS--2	-0.3852	-0.3852	0.2855	-1.3491	-0.0309	0.7918
RAMOS--3	0.5334	0.5334	0.2675	1.9941	-0.0815	0.7628
RAMOS--4	-0.1702	-0.1702	0.2324	-0.7324	-0.0091	0.6958
APPLE--1	-0.1538	-0.1538	0.2027	-0.7589	-0.0098	0.6864
APPLE--2	-0.4631	-0.4631	0.2637	-1.7560	-0.0523	0.7560
APPLE--3	0.2302	0.2302	0.3338	0.6903	-0.0081	0.8474
APPLE--5	0.3497	0.3497	0.1851	1.8896	-0.0606	0.5045
APPLE--6	0.2584	0.2584	0.2349	1.1002	-0.0205	0.6924
APPLE--7	0.2685	0.2685	0.2987	0.8990	-0.0137	0.8098
APPLE--8	-0.2296	-0.2296	0.1766	-1.2999	-0.0287	0.4560
APPLE--9	-0.2850	-0.2850	0.2368	-1.1933	-0.0242	0.7025
APPLE-10	0.0527	0.0527	0.2032	0.2532	-0.0011	0.6084
APPLE-11	0.3139	0.3139	0.2150	1.4598	-0.0362	0.6330
APPLE-12	-0.3630	-0.3630	0.2415	-1.5038	-0.0382	0.7095
APPLE-13	-0.1051	-0.1051	0.2885	-0.3609	-0.0023	0.7932
APPLE-14	0.1698	0.1698	0.1969	0.8536	-0.0124	0.5711

Table 3

The Results of the Multiple Regression of Mean-Change CAT Comprehension Scores on Teacher Performance Scores - Grade 2 (N_T = 39)

The following speculation is offered to account for these results. Decoding skills are discrete, concrete skills; for example, learning initial consonants requires learning a definite number of different sounds which begin words. The child must see the grapheme and hear its accompanying phoneme. He must practice the sound, learn to use it as a cue to help him identify the word, and learn to recognize it in a variety of settings. Practice and corrective feedback will obviously be helpful. So a teaching style in which the teacher moves from child to child, explaining the decoding skill, monitoring practice, and giving corrective feedback provides the conditions that facilitate learning. A teacher who uses a variety of materials in teaching decoding may be either confusing the child or creating a situation that is too difficult to manage.

Reading comprehension on the other hand requires learning processes which can be used across a variety of materials. The more adept the child is at transferring these processes to different materials, the greater his or her reading comprehension. A variety of materials is facilitating for teaching comprehension. In this analysis none of the teaching contexts has a significant regression weight, though independent seatwork (AP-5) approaches significance. Extended individual reading probably is necessary for developing comprehension skills.

At this point in the presentation of research data the investigator usually speculates about the meaning of the data, tries to cope with the surprises such as finding that using a variety of reading materials is a negative predictor, and suggests research that ought to be done. If he or

she is explaining the results to a group of policy-makers the usual disclaimers are made, appropriate qualifications are made, and of course the need for replication or experimentation is stressed.

But the policy-makers have to form a policy at some point. Furthermore, ought not the next step in the research lead closer to the formulation of a policy? What can be said from these data relevant to policy?

Obviously, a set of teaching performance variables is highly related to measures of reading skill. Is there a discernible pattern in this array of performances? The predictors fall into two categories: how pupils are organized for instruction, how the teacher interacts with them. The presentation of an explanation to the class as a group combined with the teacher moving from pupil to pupil increases for each pupil the amount of direct instruction available from the teacher. An effective pattern of interaction is one of discrete questions, answers to questions and corrective feedback on answers given by the pupils. This combination of organization and interaction sets a style that appears useful in teaching decoding skills. In contrast, a pattern of individual reading on a variety of materials appears to be an effective strategy for developing comprehension skills. Other data in the study suggest that a pattern of explaining, discussing, and questioning is an interaction strategy that helps to develop comprehension skills.

These data and other like them in the study point to the characteristics that policy probably should have. Teaching performances are obviously significant factors. Should, however, the policy-makers attempt to specify which skills should be learned and should a level of skill be specified?

The data suggest that performances will be specific to outcomes; that is, certain performances are linked to one outcome but not to another. It is unlikely that performances common to all outcomes in all subjects across different grades will be found. Using only the data reported here the following matrix may be constructed:

Table 4

Matrix of Effective and Ineffective Teaching Performances for Grade 2 Reading Using Mean-Change Scores as the Criterion (N = 39)

Comprehension		Decoding	
E	I	E	I
WD-5	WD-1	WD-1	WD-5
R-3		AP-7	R-2
AP-5		AP-9	AP-13
		AP-11	
		AP-12	

Note that no performance appears as effective for both comprehension and decoding; note also that a performance effective for comprehension is ineffective for decoding (WD-5) and a performance effective for decoding is ineffective for comprehension (WD-1).

Assume that results like these are obtained on a replication of the study. The possibilities and limits of a policy based on these data seem clear. The policy should specify no more than that training specific to outcomes be given, that ineffective teaching performances be identified, and that evidence be provided that prospective teachers are evaluated in

terms of multiple criteria of effectiveness. Such a policy is consistent with the available data. It is not necessary that the specific performances to be acquired be spelled out in a policy. But it is clear that two domains of performance are critical: how instruction is organized and managed, and the specific patterns of interaction used in instruction. These and similar data may be used to construct guidelines for training and evaluation.

It is unlikely that research will produce a list of specific performances that can be demonstrated as unequivocally necessary for effective teaching or which must be avoided. But research on teaching effectiveness is beginning to reveal the general characteristics of competence. I suggest that if we move back a step from the available data, it is possible to see what these characteristics are likely to be. These characteristics, abstracted from the specific research, can provide a basis for intelligent policies.

One other question needs attention. Is it reasonable to expect that standards of performance can be set? The data from Phase II suggest two answers to this question. There are any number of specific performances which appear as negative predictors of learning. The more these performances occur in teaching styles, the more likely it is that children will learn less. A teacher whose style is characterized by several or all of these performances is likely to be considerably less effective than other teachers. Using the regression equations generated in the Phase II data analysis, it is possible to make an estimate of the limits beyond which a teacher is likely to be very ineffective. Such estimates might be used

as guidelines for evaluating the skills of teachers in training. Again moving back a step from the research, it is possible to see that a policy that requires evaluation against a minimum standard of performance in practicums, internships and practice teaching has merit even though the licensing agency may be in no position to establish what these minimum standards should be. Guidelines for provisional standards appear to be a real possibility.

The other answer to this question can only be touched on here. The study found relations between different patterns of improvement in learning and distinct patterns of teaching performances. Some teachers are highly effective with most of their students; others with only some. The latter with training might be equally effective with most of their students. It is likely that teachers in the elementary school are differentially effective both in what they teach and whom they teach. The notion of the omnicompetent teacher was not substantiated in the Phase II research. It seems unlikely that setting common standards for all elementary teachers is going to be a workable policy.

In summary: if one moves back a step from the details of the research, there are three policy issues on which Phase II seems to give some guidance. First, should common standards be set for all elementary teachers? The tentative answer is no. Requirements for performances to be acquired will probably differ by subject taught, specific outcomes within each subject, and by grade taught.

Second, effective teaching probably requires using teaching performances found to be effective and eliminating those found to be ineffective. It is not sufficient only to acquire effective performances or only to eliminate ineffective ones. Evaluation of effectiveness should assess degrees of effectiveness and ineffectiveness. Assessments of prospective teachers should measure both many criteria of effectiveness and ineffectiveness and their relative degrees. It seems highly unlikely that a single summative evaluation such as a grade in practice teaching will adequately represent a teacher's competence. If this is true, then it seems wise to begin to lay out the categories of information that will be needed to implement a licensing policy.

Third, it seems possible that broad guidelines defining minimum standards of competence can be developed. These guidelines can at least suggest the major categories of effective and ineffective teaching performance and the profiles of greater or less effectiveness. Such profiles can be used in evaluating prospective teachers for certification. What I have in mind is something like the Atlas of the Minnesota Multiphasic which was a book of different profiles on the Multiphasic's scales sorted by clinical entities. A similar atlas of teaching performances sorted by their relation to degrees of pupil achievement could be used clinically to make judgments about the probable effectiveness of prospective teachers.

The implication of this proposal is that minimum standards for each type of performance is probably unwise. The performances are too correlated to treat them as independent entities. A series of stepwise regressions performed in the Phase II analysis suggest that it will be possible to identify a variety of more or less effective patterns of teaching performance.

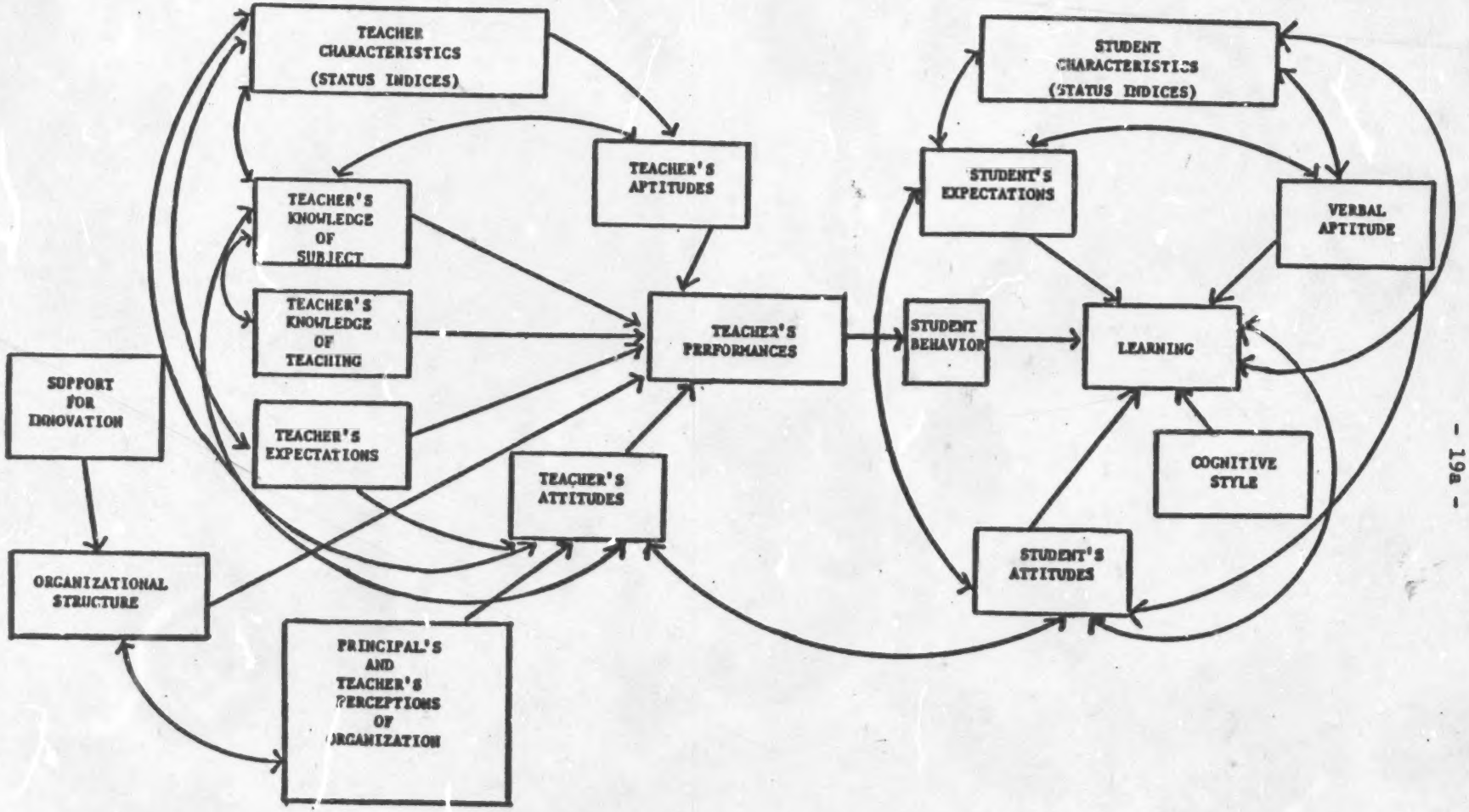

Figure 1. A STRUCTURAL MODEL OF THE DOMAIN OF VARIABLES INFLUENCING TEACHING PERFORMANCES AND CHILDREN'S LEARNING

Copyright © 1975 Educational Testing Service. All Rights Reserved.

Phase II was designed to provide data related to other policy issues. The discussion to this point leads to the conclusion that teaching performances do make significant and probably larger differences in children's learning than many have believed. But it is also important to know what the antecedents of effective and ineffective performances might be. This information will help us to estimate whether selection procedures, special training, or acquisition of knowledge are likely to be useful in producing competent teachers.

Figure 1 presents the structural model that was developed to test the relative effects of different antecedents on teaching performance. It has three major parts. To the right are variables defining learning and differences among students. The latter are predicted to have a direct effect on learning. Curved lines represent covariation of the variables; straight lines represent covariation but also predicted "causal" relations; for example, student aptitude is predicted to have a direct effect on learning but learning is not predicted to have a direct effect on aptitude. Curved lines represent zero-order correlations ($\underline{r}$); straight lines represent partial correlations ($r_{1 \cdot 234...}$).

In the center approximately are the teacher performance and student behavior variables. The prediction is that teacher performance determines student behavior, what the child does during class, and this behavior determines learning. We recognize that student behavior in particular instances does have an effect on what the teacher does, but the model represents the more general case. Particularly in elementary schools teacher performance is likely to be the controlling event.

To the left and in back of the teacher performance variables are two kinds of variables: teacher variables such as teacher characteristics (background indices), teacher aptitude, teacher knowledge, teacher attitudes, and teacher expectations; and variables describing the teaching environment. These variables are treated as antecedents of the performance variables.

Other investigators probably will draw other models. These differences are of little import at this point. Data and theory do not so strongly support one prediction that another is not valid. The model is not an induction from existing data nor from theory, though its conception has been influenced by both. Rather it is a model of how variables relate to each other by definition. Aptitude, for example, is a general trait, an information-processing characteristic of a person which affects how the person responds in many different situations. In this sense it is antecedent to observable behavior.

I emphasize that the model is a macromodel, a picture of the general flow of events. Particular instances may not be described by it. In the simplest terms it says that how the school is organized and how the principal administers the school will affect how teachers teach; the teachers' aptitudes, and knowledge, and expectations and attitudes will affect how they teach. It says that what teachers do will affect how children act in class, how they attend, how productive they are, and that in turn how a child acts in class will affect how he or she learns. But the child's aptitudes, expectations, attitudes, and his or her background will also affect how he or she learns.

The conclusive value of this model is the various relations for which it provides a test, relations of significance to policy issues. The most significant relation, that of teaching performance to learning, has been briefly described. This study consistently found significant relations between teaching performances and learning. The set of student variables surrounding "learning" in the diagram by themselves account for about 20 to 30 percent of the variance in pupils' scores. Recall, however, that in this study a pretest was used. The correlation between pretest and posttest scores is about .90. The pretest scores are correlated substantially with the student-variables scores, and account for about 80 percent of the variance in posttest scores. Almost all of the remaining variance is usually accounted for by the teaching-performance variables. These data support the earlier statement that policies should focus on teacher performance.

Against this background, we now ask, what determines differences in teaching performance? To answer this question and to illustrate how policy issues may be illuminated by this analysis, I will use only the teaching performance variables found to be significant in the regression analyses described previously. The technique of path analysis was used to estimate the strength of the relations portrayed in Figure 1. The figures reported in Table 5 are path coefficients which are like partial correlation coefficients. We have used .25 to consider a path coefficient worth nothing.

Each of the numbers in Table 5 represents the amount of covariation between two variables when the covariation of all other variables with the dependent variable has been partialled out. The path coefficient represents how much of the variation in the dependent variable is attributable to the antecedent alone.

TEACHER VARIABLES	TEACHING PERFORMANCES: POSITIVE PREDICTORS				NEGATIVE PREDICTORS			
	WD-1 INST. TIME	AP-7 CLASS	AP-9 QUESTIONS	AP-12 FEEDBACK	WD-5 VAR. MATR'LS	R-2 BREADTH	AP-11 LOCATION	AP-13 DISCIPLINE
ATTITUDES:								
Aspirations	-.0232	.4894	-.1904	.0024	.2226	.0905	.3091	.3280
Satisfaction	.2015	-.2098	-.0888	.3791	-.1422	-.2073	.0970	-.2069
Perc. Stds.	-.3815	-.3976	-.1350	-.0142	.5333	-.0069	-.7822	-.0070
KNOWLEDGE:								
Methodology	.3598	-.1624	-.2933	-.2100	-.3583	-.0136	.4409	-.1208
Subject	-.2731	-.4225	-.4521	.1712	.0557	.0506	.0225	.0108
APTITUDE:								
Verbal Fluency	.1575	.0964	.1327	.4879	.1089	.0908	-.2836	.2492
Memory	-.2345	.0108	.1293	.3337	-.0378	.1224	-.3664	-.2188
Reasoning	.0282	.1117	.6587	-.0872	-.7753	.2196	.2619	-.2106
Flexibility	.2520	.3423	-.2296	-.2531	.7759	-.0894	.3421	.3022
Cognitive Style	-.0266	-.1741	-.6043	-.4217	-.3752	-.5415	.1771	-.3164

Table 5

Path Coefficients of Teacher Variables to Significant Teaching Performances
Grade 2 Reading (N_T = 39)

Table 6 presents a boxscore of the significant relations found (.25 or better). The sign is noted when the path coefficient was negative. There are three summary columns, one for positive predictors, one for negative predictors and one for both. In the right-hand column is the ratio of the number of significant relations found to the total possible.

Recall the questions to which these data are relevant: do attitudes, knowledge and aptitude determine performance? The summary ratio indicates that they do. But the specific relations are complex.

In the attitude domain, for example, six of the eight significant relations were found with an attitude variable that measures how teachers perceive students. The high end of this scale indicates that the teacher perceives his or her pupils as having a better educational background, as coming from a higher socioeconomic background, and as being easier to teach. In the knowledge domain the results are mixed. The relation of knowledge of subject to performance should be accepted cautiously. The measure of knowledge was a measure of the teachers' decoding skills. The relation obtained, however, indicates that having phonic skills is inversely related to asking questions. Three of the four relations of this kind of knowledge to the positive predictors are negative. The aptitude domain produced the largest number of relations. The cognitive style measure had a pattern of consistent relations with how the teachers taught. Note also that all the aptitude variables had significant relations with the feedback and location performance variables.

I resist here the temptation to speculate about what all these relations mean. The final report of this project contains such speculations. I prefer now to emphasize the implications of these data for policy-making.

TEACHER VARIABLES	POSITIVE PREDICTORS: WD-1 INST. TIME	AP-7 CLASS	AP-9 QUESTIONS	AP-12 FEEDBACK	T_P +	T_P −	NEGATIVE PREDICTORS: WD-5 VAR.MATR'LS	R-2 BREADTH	AP-11 LOCATION	AP-13 DISCIPLINE	T_N +	T_N −	T +	T −	HITS TOTAL
ATTITUDES:															
Aspirations		x			1	0			x		2	0	3	0	8/24
Satisfaction					1	0					0	0	1	0	
Perc. Stds.	x (-)	x (-)		x	0	2	x		x (-)	x	1	1	1	3	
KNOWLEDGE:															
Methodology	x		x (-)		1	1	x (-)		x		1	1	2	2	7/16
Subject	x (-)	x (-)	x (-)		0	3					0	0	0	3	
APTITUDE:															
Verbal Fluency				x	1	0			x (-)	x	1	1	2	1	20/40
Memory				x	1	0			x (-)		0	1	1	1	
Reasoning	x	x	x		1	0	x (-)	x	x		2	1	3	1	
Flexibility				x (-)	2	1	x		x	x	3	0	5	1	
Cognitive Style			x (-)(1)	x (-)(1)	0	2	x (-)(1)	x (-)		x (-)(1)	0	3(1)	0	5	

(1) Negative sign indicates that field-dependence is associated with amount of questioning.

Table 6

Summary of Significant Path Coefficients from Teacher Variables to Significant Teaching Performances Grade 2 Reading (N_T = 39)

You will recall that analyses of the Coleman data found significant correlations between a measure of teacher aptitude and pupils' achievement. This measure of aptitude was also used in this study and is a component of the verbal fluency factor. The correlations of this factor with mean change scores for second-grade reading range from -.03 to .17. As the data in Table 5 indicate, the verbal fluency factor is related only to one positive predictor, giving feedback. The conclusion to be drawn is that if verbal aptitude measures are used to select prospective teachers, the measure will not necessarily select candidates who will be competent teachers. The argument for using a verbal aptitude measure has to be made on other grounds, such as, if prospective teachers have a higher level of verbal aptitude, they will be able to learn more complex and difficult content, the learning of which is probably a necessary but not a sufficient condition for being a competent teacher.

The cognitive style measure, however, and the flexibility measure are fairly good predictors of performance. These data are consistent with the hypothesis that aptitude acts as an information-processing variable which influences performance. But the data do not automatically lead to the recommendation that measures of cognitive style and flexibility should be used in a program to select teacher candidates. But they do suggest that measures of such characteristics will be useful in designing specific training programs for each candidate.

The logic of this recommendation is this. The data suggest relatively strong relations between aptitude and performance. If we assume that candidates are likely to be influenced by these aptitude factors in how they teach, a training program should counteract or enhance this tendency.

Each program, therefore, must be tailored to some degree to the aptitude characteristics of the teacher candidates. Policies which set standards for training programs should recommend that these programs be adapted to the differences in aptitudes among candidates.

The results in the knowledge domain are mixed, but knowledge of methodology does not appear to be strongly related to effectiveness or to ineffectiveness. A policy that allows for the tentativeness with which these data should be held would be one requiring that knowledge of methodology be related specifically to performance.

The attitude domain produced the smallest number of relations, and these were associated with the measure of the teachers' perceptions of students. A fair conclusion from the data is that teachers with higher scores on this measure spend less time in direct instruction, teach the class-as-a-whole less, and spend less time moving around the class, probably supervising instruction from their desk. On the other hand, teachers who perceive their classes as more difficult to teach, as coming from lower socioeconomic homes, and as having a poorer educational background spend more time instructing, ask more questions, teach the class-as-a-whole more, and move around among the pupils.

It is difficult to see what policy might emerge from a consideration of these data. Obviously teachers are adapting their styles to their perceptions of their students. But the teaching performances were found to be effective irrespective of the pupils' social class background or previous learning. Perhaps policy should require that prospective teachers both teach children of different backgrounds and be evaluated on whether or not they teach children of different social class background in ways that are effective irrespective of the pupils' background.

These results in these domains do indicate that setting standards for training and licensing will not be simple. Obviously research is needed both to substantiate the results presented here and to elucidate them. One obvious path for research is to explore the conditions required to help prospective teachers acquire teaching performances found to be effective. How trainable are teacher candidates? Is their learning so dependent on their aptitude characteristics that some of them will find learning some teaching performances too difficult? Can training conditions be made more effective if they are adapted to these aptitude characteristics? Will such training be more expensive? How does knowledge of teaching methodology or of what is to be taught affect teaching style?

Answers to such questions will help refine policy. But the results of Phase II, in my judgment, indicate the shape that policies should take. It seems reasonable that specific performance training should be required; and, that such training should be different for individuals who have different aptitude characteristics. The evaluation of prospective teachers should be more concrete and diverse. Candidates should be evaluated in terms of their effectiveness with different kinds of pupils.

If research is to influence policy, such implications have to be drawn at each stage of the research. Furthermore, it is my belief that it is not necessary to wait until the research is complete before making policy. Two reasons support this belief: the research would be poor research if it were complete; and in other areas of human activity we function on the basis of what we now know, and revise and modify policies as new knowledge develops.

A narrow legalistic conception of policy-making is to be avoided. Policies can be broad; they can provide guidelines and recommendations for action. It is premature to require that teachers, for example, learn certain kinds of teaching performances. But there is sufficient evidence, in my judgment, to redress the balance between theory and practice in teacher training. It is premature to specify admission procedures but requiring that more information be gathered on prospective teachers and be used in planning their programs is sensible in the light of the data.

As I stated earlier in this paper, other kinds of research are needed, such as cost analyses, before policies can be fully implemented. But when programs are being revised or new programs planned, the available data can be used to develop policies for making these changes. If policy-making is viewed as a developing and experimenting process, it is sensible to use the research from Phase II to make provisional policies whose effects can be studied. If scope is left for the judgment and inventiveness of those who will implement the policy, implementation will come to be seen as an integral part of the process of making policy.

You may not agree with the implications for policy that I have drawn from this small sample of the Phase II research. I hope that you have better ideas than those that I have offered. My goal was to illustrate a process, one in which the researcher steps back from the concrete details of the research, and asks what does it mean for policy? It is our responsibility as researchers to speculate about the implications of our research and to evaluate the tenuousness of inferences which can be drawn from it. In the end, the policy-makers will decide what is to be done.